

REMARKS

Applicant gratefully acknowledges the examiner's indication that the previous rejection has been overcome. Moreover, claims 3-9 remain allowable.

Claims 1, 2 and 10 stand rejected 35 U.S.C. §102(e) as being anticipated by Hamamura. There is some difficulty understanding the Hamamura patent. However, it does not appear to be directed to a method for segmenting compound words in an unrestricted natural-language input. Rather, Hamamura appears to be directed to character and word recognition in a restricted word environment. For example, it appears to repeatedly use the example of inputting a city name and providing a word dictionary with a list of permitted city names. Moreover, Hamamura does not appear to be segmenting compound words but rather recognizing words each separated by a space. Character spacing plays a role in the word detection system of Hamamura as set forth at column 18, lines 59-column 19, line 18.

The methods of claims 1, 2 and 10 operate on a plurality of characters. Indeed, claim 1 specifies, "consisting of a plurality of characters." Characters are defined in the application as the symbols that make up the natural-language input. These may be letters or punctuation marks or the like. The invention is for segmenting compound words in other words, those not already separated by white space. Whereas Applicant's invention teaches how to segment a word such as "geschwerkschaftsorganisation", Hamamura seeks to recognize a city name such as "st lin". For Hamamura, white space plays a role in identifying the words.

Applicant's invention further requires "constructing a set of probabilistic breakpoints in the natural-language input." The examiner refers to Hamamura at column 14, lines 42-45 which describes probabilities for character or word recognition. The number of characters in the word are known. There is no discussion of segmenting the word into linkable components. To the contrary, such segmenting is specifically disclaimed by Hamamura at column 14, lines 60-64. "In a language (such as English) in which a word break occurs, the methods described in the foregoing chapters 1-4 assume that a word is always identified correctly. If the number of characters is changed while this assumption is not met, these methods cannot be used."

Thus we refer to a second section identified by the examiner at column 21, lines 23-43. Again, segmenting a compound word is not disclosed or suggested by Hamamura. The character spacing is critical. Hamamura states at column 21, lines 61-63, "If no character spacing is provided while this assumption is not met, the above methods cannot be used." Thus, these portions of Hamamura are not used for segmenting a word defined by spaces at either end.

We refer now to the third section of Hamamura identified by the examiner column 32, lines 66-column 33, line 18. This portion refers to "The result of character recognition by conditioning characteristics of the characters and character spacing of the words contained in a word dictionary that stores candidates of the characteristics of a word to be recognized and character spacing of the word." Column 33, lines 6-10. Rather than segmenting a compound word, Hamamura attempts to identify words separated by spaces. This fourth embodiment of Hamamura is designed to provide character delimiting useful for kanji such as described at column 26, lines 26-35.

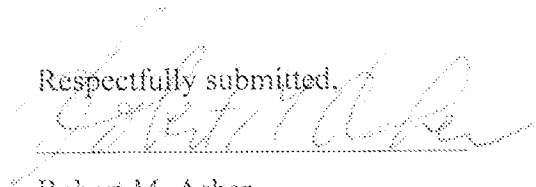
The Applicant's invention further requires traversal of substrings of the natural-language input that have been delimited by the breakpoints. By way of example the invention describes traversal of breakpoints at page 20, paragraph 005. Such traversal of potential breakpoints is not found by Applicant in Hamamura.

Claim 10 is more explicit in specifically identifying the order of traversing breakpoints. Claim 10 specifies that the order is determined by probabilities obtained in the probabilistic breakpoint analysis. The examiner refers to the first embodiment as described at column 8, line 53-column 9, line 13. The first embodiment is providing word recognition on a word having a known number of characters as set forth above, at column 14, lines 62-65, "If the number of characters is changed while this assumption is not met, these methods cannot be used." Hamamura gives an example of a restricted choice of cities giving five city names at column 8, lines 62-65. A probability is determined for each of the potential city names. This embodiment does not describe determining breakpoints in the word through a probabilistic analysis and then traversing the substrings delimited by the breakpoints. Far from it, this embodiment merely tries to recognize a word of known length. Note that each of the five city names in the lexicon

has four letters. In the example referred to by the examiner, character recognition produced the result "haia", but the actual written character was "mair". The restricted lexicon was helpful in overcoming the limitations of simply using character recognition to identify the city. Indeed, Applicants find Hamamura inapplicable to their claimed method for segmenting compound words in a unrestricted natural-language input.

For all the foregoing reasons, Applicant submits that all claims pending in the present application are allowable over the art of record and early notice to that effect is respectfully solicited.

Respectfully submitted,



Robert M. Asher
Reg. No. 30,445
Bromberg & Sunstein LLP
125 Summer Street, 11th Floor
Boston, MA 02110-1618
(617) 443-9292
Attorney for Applicant

02639/00A36 581637.1